Retrieving Pressure-Temperature and Water Vapour Profiles in Earth's Atmosphere from INSAT 3DR data using Machine Learning

Fida Salim, Soumik Bhattacharyya

Abstract

Current methods of predicting sea surface temperature (SST) using satellite data involves certain physical assumptions subjected to initial conditions and boundary. In our project, we want to use machine learning to model the relation between brightness temperature (BT), which can be directly obtained from real time satellite data, with actual pressure-temperature (PT) and water vapour (wv) profiles. As part of data collection and pre-processing, we obtained 25000 PT and wv profiles over Indian Ocean from ECMWF library and retrieved corresponding BTs for each using a forward radiative transfer model, called RTTOV. We tried to reduce the dimension of the PT and wv profiles using PCA and regressor models. We also tried an initial attempt at the modelling using Random Forest and Neural Network, Autoencoder. The average percentage error in predicting temperature using Autoencoder and Random Forest with two different hyperparameter sets are respectively 1.65%, 1.30% and 1.05%.

1 Introduction

The temperature of the water at the ocean's surface, or sea surface temperature (SST), significantly impacts climate and weather. The change in SST can cause alterations in rainfall patterns and sometimes lead to the development of tropical cyclones. In order to retrieve SST, atmospheric scientists usually rely upon satellite observations and in-situ instruments like buoys, ships and ocean reference stations. However, taking regular in-situ measurements from all over the world isn't feasible and satellite based observations only provide the brightness temperatures at the frequencies in which the satellite operates. The actual temperature is then found using time-dependant regression models, which are based on assumptions subjected to boundary and initial conditions.

However, the opposite, retrieving brightness temperature from actual temperature, is quite easily possible using forward radiative transfer models. RTTOV, developed by ECMWF, is such an example of a fast one-dimensional radiative transfer forward model. It determines radiances and brightness temperatures (BT) from an atmospheric temperature profile, variable gas concentrations, and cloud and surface characteristics. If it is assumed that certain window regions of the infrared and microwave frequency spectrum have little interaction with the atmosphere, the brightness temperature obtained from RTTOV and the surface temperature can be comparable.

1.1 Project Problem Statement

The default input profiles for RTTOV are the traditional temperature and water vapor profiles, which will provide the brightness temperature. This model is well-appreciated and acclaimed. But, what struck our mind is, whether we can obtain the water-vapor profile and pressure-temperature (PT) profile merely from certain brightness temperature values. Our goal is to create a machine-learning model to learn and build PT and water-vapor profiles from the input of brightness temperature values.

The implications of retrieving these atmospheric profiles with a good confidence level can help contribute to weather forecasting and compensate for the lost information due to the scarcity of in-situ observations. A machine-learning approach to this problem is carried out as it involves handling a large amount of the data-sets, which could be trained to find a one-one relation between input and output dataset. Later, the model can be used on real time data of INSAT 3DR, an Indian geostationary weather satellite.

1.2 Literature survey

The relationship between water vapor and sea surface temperature has long been a topic of interest in atmospheric science. *Stephens, 1990* provides a literature review in this field, along with his own analysis of satellite observations to illustrate the spatial and temporal patterns of this relationship. Machine learning techniques have been adopted previously in order to forecast SST from previous observations. However, errors from previous iterations propagate in these time series regression models, as is experienced by *Sarkar et al., 2020*. They used a convolutional neural network (CNN) and a recurrent neural network (RNN) to analyze satellite data and predict sea surface temperatures. Some other works related to our domain involve *Tripathy et al., 2006* (ANN to predict SST anomalies in Indian Ocean); *Xiao et al., 2019* (Long short-term memory (LSTM) and AdaBoost to predict SST) etc.

2 Baseline Algorithm

2.1 Non-Linear Regression Models

The relation between atmospheric pressure and temperature, as well as water vapor and pressure, is non-linear. We have the requirement to parameterize the P-T (w-v) profiles so that a one-one relation between their corresponding brightness temperature (BT) can be later inferred. Specific non-linear regressor models were used to capture the complex relationships between pressure and temperature (water vapor).
The regression models used are: (i) Gaussian Process Regression (`GPy`), (ii) k-NN regression (`sklearn.neighbors.KNeighborsRegressor`), (iii) Support Vector Regression (`sklearn.svm.SVR`), (iv) Decision Tree Regression (`sklearn.tree.DecisionTreeRegressor`) and (v) Random Forest Regression (`sklearn.ensemble.RandomForestRegressor`)

2.2 Dimensional Reduction Models

Our data, the P-T (w-v) profiles, are available for 25000 observations with 137 fixed height points. So, we have a dimension of [25000,137] for our model to learn on. Thus dimension reduction methods are used to reduce the number of features in the dataset. **PCA** (`sklearn.decomposition.PCA`), the feature extraction method, is used for data compression, allowing large datasets to be stored and analyzed more efficiently.

3 Experiment

3.1 DataSet

- 25000 pressure-temperature and water vapor profiles from the European Centre for Medium-Range Weather Forecasts (ECMWF)[Data].
- The RTTOV forward model was used with P-T and w-v profiles to get corresponding brightness temperatures at INSAT-3DR sounder channel frequencies [Data].
- The INSAT 3DR imager data will be used in future for implementing our ML model.

3.2 Non-Linear Regression

We wanted to see whether a non-linear regression model can be utilized to parametrize the pressure-temperature relation for each profile. By following this path, we had to realize that each profile can be mathematically represented so that a one-one correlation with Brightness Temperature can be easily established. Various regression models were used to choose the best fit, avoiding over-fitting models and using validation sets to determine the hyperparameters, see Table 1 and Table 2. One of the profile among 25000 were randomly chosen, and parametrization was done on that profile. 20% of the data was used as test data, 20% was used as a validation set, and 60% was used as the training set. Using the Validation set, the optimum value for hyperparameters was obtained for each model, and R2 score and MSE was determined. Regression line for each model is illustrated in Figure 1

Model	Hyperparameters	R2 Score	MSE	Train Score	Test Score
GPR	α = 10, length scale = 2	0.999	0.689	0.999	0.999
k-NNr	k = 3	0.995	3.732	0.981	0.999
DTr	max_depth = 15, min_sample_split = 2	0.996	3.296	1	0.996
RFr	n_estimators = 100, max_depth = 15	0.998	1.195	0.999	0.998
SVR	C = 1000. ϵ = 0.01	0.99	7.572	0.988	0.99

Table 1: Comparing Models for P-T profile

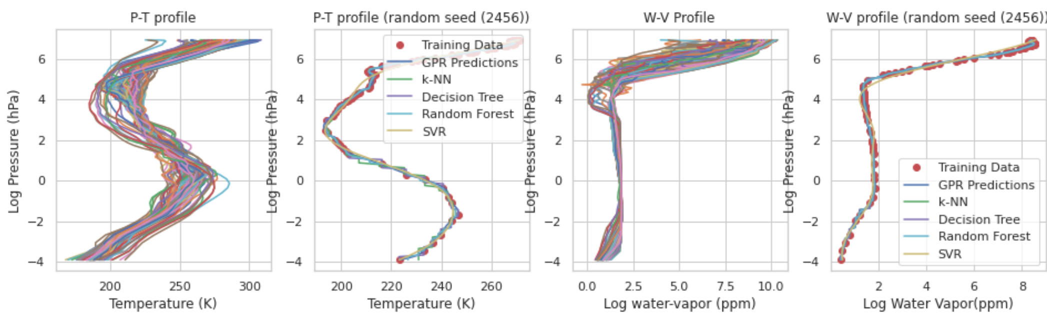
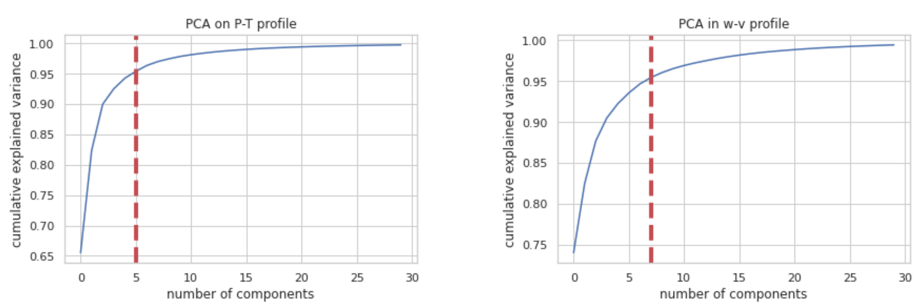


Figure 1: Regression model in P-T and W-V profile

Model	Hyperparameters	R2 Score	MSE	Train Score	Test Score
GPR	Alpha = 1, length scale = 0.1	1	0.005	0.999	1
k-NNr	k = 5	0.997	0.011	1	0.999
DTr	max_depth = 15, min_sample_split = 2	0.999	0.011	1	0.999
RFr	n_estimators = 50, max_depth = 15	0.999	0.009	1	0.999
SVR	C = 1000. epsilon = 0.01	0.989	0.108	0.991	0.989

Table 2: Comparing models for W-V profile.

3.3 PCA



(a) Cumulative Explained Variance for P-T profile

(b) Cumulative Explained Variance for W-V profile

Figure 2: PCA applied on the profiles

As depicted in Figure 2, PCA was used as a feature extraction method on P-T and W-V profiles. The cumulative explained variance (CEV) is a useful metric for assessing how well PCA reduces the dataset's dimensionality. PCA for P-T and W-V profiles were used with 30 components and obtained that five or above principle components in P-T profile give a CEV of more than 95% and seven or more principle in the W-V profile give CEV of more than 95%. With the help of this information, 9 components of each was used in ANN for modelling temperature.

3.4 Modelling Temperature from Brightness Temperature

In order to model the actual temperature from brightness temperatures, we employed two different initial approaches involving a neural network model, the Autoencoder (`tensorflow.keras`), and a random forest regressor (`sklearn.ensemble.RandomForestRegressor`). PCA has been applied on both P-T and w-v profiles to reduce the output dimension of the Autoencoder to 9 components each. After 50 epochs, our model was able to predict the temperature of the test dataset with an average percentage error of 1.65%. However, the random forest regressor outperformed the Autoencoder in terms of prediction accuracy, as evidenced by the following table, which shows the percentage error in predicting temperatures for two different forests with varying numbers of trees and nodes per tree [1].

No of Trees (N)	Nodes per tree (d)	Time Taken	Model Score (%)	Error in Temp. (%)
100	10	30 min	61.58	1.30
1000	100	6 hours 40 min	70.55	1.05

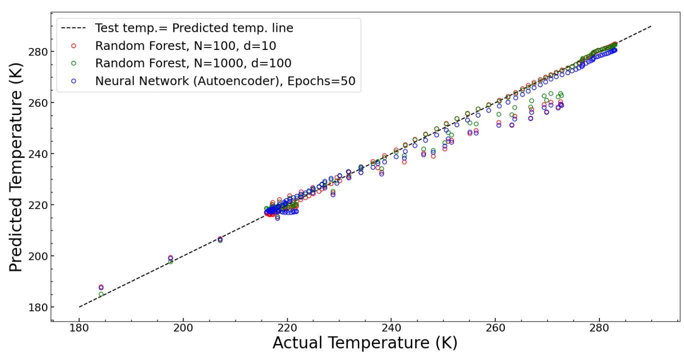


Figure 3: For a randomly selected profile, the predicted temperature using model is plotted against the actual temperature. The black dashed line shows the x=y line for reference.

4 Future Plans

- To develop our ANN and Random Forest based models to increase accuracy.
- To use MLP as a model to predict atmospheric profiles from BT.
- To use the model on INSAT-3DR data

5 References

[1] G. L. Stephens. (2000) On the Relationship between Water Vapor over the Oceans and Sea Surface Temperature. *J. Climate*, 13(4): 657-666.

[2] P. P. Sarkar, P. Janardhan & P. Roy. Prediction of sea surface temperatures using deep learning neural networks. *Neural Computing and Applications*, 31: 3665-3674.

[3] K. C. Tripathi, M. L. Das & A. K. Sahai. Predictability of sea surface temperature anomalies in the Indian Ocean using artificial neural networks. *Journal of Marine Systems*, 63(3-4): 125-140.

[4] C. Xiao, N. Chen, C. Hu, K. Wang, J. Gong & Z. Chen. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sensing*, 11(17): 2041.

[1] The python code for our experiments can be obtained in this github link.